

**September 10, 2009**

**RE: Portland Harbor RI/FS Bioaccumulation Modeling Report, Dated July 21, 2009**

**General comments:**

Given the different sources at the site (dioxins / furans, DDX compounds and different PCB sources), a more focused BSAF effort should be done to evaluate the bioaccumulation characteristics of these different sections of the river, or the food web model should be re-designed to apply to these areas. Applying a food web model focused on averages in a river with significant and distinct sources (in spatial scale and characteristic) has a high degree of uncertainty which will translate into PRGs development.

**Biota/Sediment Relationships:**

- The focus on BSARs in such a variable environment should be discussed. This is a sediment area with a lot of different sources. The requirement here was that a pretty significant relationship between all stations sampled had to occur in order to proceed with modeling. In fact, if there was only a few areas with detects and relationships, this ultimately resulted in an “insignificant” site-wide relationship. In addition, it is unclear how analytical differences between Round 1 and Rounds 2 and 3 played out in these models. However, it is clear that there were “significant analytical differences between the two datasets.” (page 38). For these reasons, BSAFs should have been more fully explored for all chemicals including those modeled in the food web model. DEQ has had shown good relationships in the past using these techniques.
- It was my understanding that EPA requested BSAFs to be developed for chemicals that were also modeled in the food web model. This was not provided and represents a significant data gap.
- Round 1 and Round 3 smallmouth bass data need to be treated differently in the models based on compositing differences. While DEQ showed good BSAF relationships with Round 1, Round 3 relationships are even stronger because compositing occurred over more localized areas corresponding to exposure areas. It is unclear how this was handled in the report

**Food Web Model:**

- There is still no true evaluation of the data fit relative to model performance that considers the actual range of the data relative to predictions. The model performance (SPAF) is based on standard errors on the mean instead of standard deviation. Figures that show model performance relative to empirical tissue are presented on a log scale.
- Multiple parameters were calibrated that should instead be represented by the variability in the environment. The biggest example of this was water concentration. The result is still a highly calibrated model. Uncertainty should be evaluated by running the empirical data through an

uncalibrated version of the Gobas model. After all, according to Gobas, “there is rarely a need for model calibration as long as the model is used within its application domain”. Model calculations should be checked for consistency with available empirical data to gain confidence in the model, but this should be done by instead comparing model outcomes with independent data (data not used in the construction of the model).

- Additional metabolism parameters were added for invertebrates and fish for compounds which likely are not metabolized to a significant degree. These rates were ultimately calibrated to get a best fit.
- A focus on the average dilutes out large elevations in concentrations, which occur in several areas of the river. The most apparent based on the smallmouth bass data is around river mile 7 west for dioxins and furans, but others for PCBs also exist (e.g. RM 2E, 11E). There is some discussion of this relative to DDX compounds, but not to dioxin and furans. Dioxin TEQ in the 7W area of the site are significantly elevated relative to the rest of the harbor, however this neither shows up in the model nor is it discussed in the BSAF/R section of the report. For example, smallmouth bass sample SB7W had a fish TEQ of 70 pg/g, a bird TEQ of 262 pg/g, and a mammalian TEQ of 57 pg/g. The drivers of TEQ risk in these fish were 2,3,7,8-Tetrachlorodibenzofuran (110 pg/g), 2,3,4,7,8-Pentachlorodibenzofuran (108 pg/g), 1,2,3,7,8-Pentachlorodibenzofuran, and 1,2,3,4,7,8-Hexachlorodibenzofuran, and 1,2,3,6,7,8-Hexachlorodibenzofuran (bird example TEQ's in parenthesis). Also a significant contributor was PCB 77 (24 pg/g). A graph of showing contribution to TEQ for dioxin like PCBs and dioxins and furans is included to illustrate these trends. A bigger question is whether the PRGs developed are protective of the different sections of the harbor.
- There is no evaluation of how the model predicts relative to species other than sculpin and smallmouth bass. It doesn't appear to perform well for carp, for example. Round 3 carp showed significantly elevated concentrations relative to Round 1, with one composite sample (3 miles) at 25,090 ppb Total PCBs and another 615 ppb Total DDX. These samples represent huge outliers to the model, but yet represent large composite areas.
- The model was calibrated on Total PCBs instead individual congeners. This is a deficiency in that it wasn't calibrated on a congener you could consistently track in all media: sediment, tissue and water. Dioxin like PCB congeners, for example, change throughout the harbor (see figures 1-4 attached). The evaluation of the fit of specific congeners is uncertain given “calibration” using metabolic rates was conducted.
- Treatment and calculation of water dissolved water concentration data was a previous comment, and the same methodology was used in this report (Appendix E, Page 5). The model should be evaluated without the calculation made on the empirical data to evaluate the uncertainty in this methodology. In addition, the use of the water data in the model is needs further review. There is a high degree of uncertainty in using the average of transect data, and further “weighting” this data by using estimates of contribution to total flow (see specific comment). I see this as one of the bigger issues with the food web model. Water concentrations were averaged and adjusted for contribution by hydrograph period (e.g. storm water and low flow contributed 78% of total to “concentration” while high flow only 28%). It is

inappropriate to adjust water concentration in this manner, but rather distributions should be used.

- Not all tissue data collected by the LWG was used in the modeling effort (see Figure 4-1 through 4-7), and there are questions about which water samples were used and how they match up with the tissue included or omitted. Submittal of the database used for the modeling effort is necessary to answer these questions. I would also like to request an electronic copy of the food web model.

### **Specific Comments:**

Table 2-1: BSAFs should have been calculated for food web modeled chemicals. This would have included smallmouth bass to chemicals like PCBs, dioxins and furans and organochlorine pesticides. Based on my examination of the data, some of the best relationships are between the Round 2 smallmouth bass data and sediment chemistry.

In addition, it is unclear why BSAF methodology was used for all other fish with the exception of smallmouth bass, which used a BSAR approach.

BSAFs were not completed for the mussels based on a footnote that states “No BSAF or BSAR was developed for these species because their tissue contamination is expected to be driven by water exposure (rather than sediment exposure)”. This statement is not supported; further, BSARs were calculated for clams, which feed in a similar manner.

Section 3.2, Page 12: No mention of Total TEQ – the combination of PCB TEQ and dioxin and furan TEQ. All dioxin like activity should be summed as a total. For birds, 2,3,7,8-PCDF is not the best candidate as a surrogate congener for bird TEQ. 2,3,7,8-TCDF is a better choice. While PCB 77 is a good choice for birds, it is not a good surrogate for mammalian TEQ. PCB 118 or 126 should have been selected. See figures 2 and 3.

Section 3.3.2, Data Sets: Since datasets used to develop BSARs are not included here, we need to wait until they are presented to finish the review (Appendix H of the RI Report). This is especially true since it is stated that “any co-located pair with non-detected tissue or sediment concentrations was removed from the BSAR analysis”. Did Round 1 tissue only use the one sediment sample taken during that time to develop a relationship (e.g. see noted problems with that dataset esp. for pesticides – use of SW8081 only).

Laboratory clams should have been included in the modeling effort, especially since they represent exposure to site sediment without any confounding factor of water.

Section 3.3.3, Data Prep for Sculpin: A centroid for 0.1 miles (500ft) was used to represent the sculpin sediment exposure area. This 500 feet was circular, but should have been linear. The circular 500 ft was also “assumed to approximate the area over which the individuals in a composite sample were

collected” (Append D, Section 2.1). The area of the composite should be determined by the sampling area where they were collected, not the centroid for the composite.

Section 3.3.3, Data Prep for Smallmouth Bass: During planning for Round 3, it was decided that the bass composites should be built on a 1 mile exposure area, and that is how the fish were composited. Sediment from that 1-mile area should have been used as the exposure area. Instead, the “sediment increments were evaluated from 1-mile upstream to 1-mile downstream of the collection location” of each fish. Depending on the composite, this increased the sediment exposure area of the composite to 3 miles. analysis of the Round 3 smallmouth bass data shows very good relationships with the sediment on a 1-mile basis.

Beach sediment: It is unclear why beach sediment wasn’t also considered sediment for the purposes of determining exposure point concentrations, as it is inundated during parts of the year. Why weren’t sediment natural attenuation cores included – they represent the same depth (0-30 cm)?

Section 3.3.4, Approach for Large Home Range Species: Black Crappie were only collected in the Terminal 4 area of the site. Therefore, exposure for these species should be based on sediment from these collection locations. This is especially true since this is a breeding area for these fish.

Section 4.1.1, Model Development: Why is a difference between the slope and zero required at a  $p < 0.05$ ? This seems like a stringent requirement for model development.

It should not be assumed that an intercept greater than zero necessarily indicates significant water background water or prey exposure. There is too much noise in the analytical chemistry, depth of sediment analysis (30 cm, most of which the organism isn’t exposed to), and variability in bioavailability to attribute this phenomenon entirely to those indicated here. Based on other possible scenarios as variables in determining a relationship, I also don’t think it is defensible to state “the lack of a relationship between sediment and tissue concentrations might indicate that a medium other than sediment is the source of tissue residue (e.g. upstream or lateral loads to surface water)”.

Table 4-2, 4-3, 4-4 and 4-5: I would like to request the dataset used to develop these models, the models developed in Table 4-2, as well as the documentation to show that models could not be developed. A review of the applicability of the dataset used for development is a key part of this review.

Table 4-3, Steady State Correction Factors: It is unclear concentrations in laboratory worms were corrected for steady state conditions as described in Section 3.3.2.

Table 4-5: BSAR/F for organochlorine pesticides, dioxins and furans, and PCBs should be included. Round 3 should be considered separately from Round 1 given the difference in compositing. The combination of Round 1 and Round 3 in one linear model is likely leading to “no relationship”. BSAFs for smallmouth bass (esp. Round 3) should have been included as a part of this analysis. Based on my analysis of Round 3 bass, the 1 mile- tissue concentrations are very reflective of sediment concentrations. The bass respond very well to localized sediment concentrations.

Section 4.2, Large Homorange Species: A BSAF should have been attempted regardless of whether “at least one BSAR for a smaller-home range species could be identified for a given chemical”. It is also unclear why lamprey ammocoetes are included as “large home range species”, where if they had modeled, the relationships between sediment and tissue would be estimated by a site wide average sediment concentration and a site wide average lamprey tissue concentration.

Section 5.3.3.1, Model Calibration: Several environmental parameters such as water temperature, total suspended solids in water, dissolved OC in water, and OC content of sediment were calibrated. Values that represent site specific information should not be calibrated, but rather the variability in these values should be represented by a distribution.

It is unclear why the model was calibrated on Total PCBs, and then validated with PCB congeners. PCB congener information in sediment, tissue and water should have been the dataset used for calibration. “Total PCBs” as a group can actually represent a different congener set between media – especially water.

Section 5.3.4: Model Performance Matrix: Given the pattern of contaminant distribution, model performance needs to be evaluated beyond “average predicted” and “average empirical”. Otherwise, large peaks (source areas) can be averaged out.

Section 5.3.5, Modeling Approach: The model was run 50,000 times? I don’t think the site-specific environmental parameters should have been calibrated to this degree. How do these highly calibrated environmental parameters of the site match up with actual measurements? Again, these should be represented by distributions of the actual data. It is also unclear why you would pick the best calibrated environmental parameters for smallmouth bass (weight, lipid, water content, dietary fractions) from these runs since the model is run site-wide and small mouth bass have a home range smaller than the site.

Section 5.3.5.2.1, Chemical Concentration in Water: Water is a sensitive term in the model. The only water data used was the water collected during the seven sampling events at five transects. Transect water data is an average of water concentrations across the width and depth of the river. Fish do not use the river in this manner, and therefore it is unlikely that focusing on these water concentrations match up with fish and invertebrate exposure. This would especially be the case for sculpin, smallmouth bass, crayfish and invertebrates. Source water samples should be included. Table 5-2 should show the standard deviation, and not just the standard error, as this does not represent the range in the underlying data. We should be examining the distribution of the empirical water data.

This report does not show how the bioavailable fraction in water was calculated, which was one of the comments on the last report. It should be clear how the bioavailable fraction was calculated from the XAD samples and used in the model.

Section 5.3.5.2.2, Chemical Concentration in Sediment: This section seems to be indicating that both LWG and non-LWG sediment chemistry data were used to calculate SWACs. If correct, are all of these data Category 1?

Why was sediment concentration treated as a decision variable in the model? It was not used as a calibration parameter. The uncertainty in the average SWAC used in the model should be through calibration like any other parameter. What are the uncertainties associated with using the SWACs in Table 5-3 as current conditions? EPA recommends the use of inverse distance weighting (IDW) over the use of Thiessen polygons to develop SWACs, stating “key advantages of IDW include the ability to estimate a 95% upper confidence levels of the mean and avoiding the development of thiessen polygons that do not accurately reflect the geography of the site (e.g., extend from near shore into the middle of the navigation channel).”

What is the standard deviation around those numbers? Page 45 states “because uncertainty surrounding sediment concentrations would also apply to alternate conditions (PRGs), and distribution describing many of the uncertainties surrounding the SWAC was not included in the model calibration”. Isn’t this exactly why we need it included in the model calibration? This seems to be an important part in determining that sediment cleanup will result in the desired reduction in fish tissue. Both water and sediment can be included in the model as distributions.

Section 5.3.5.3.1.: The choice of Total PCB as the first calibration, and allowing both chemical specific and non-chemical specific parameters to vary are significant limitations to the model. Sediment should not have been held as a point estimate. Water concentration should not be calibrated but represented as empirical data.

Section 5.3.1, Calibration of Non-Chemical-Specific Parameters: Based on the text here, it appears the model may not accurately predict high carp composites, as two composite samples were discarded in order to improve the SPAF. Since there are only 15 total carp samples and they are all collected over large areas, the loss of this information is not appropriate.

Section 5.3.5.3.2, Calibration of Chemical Specific Parameters: It is not appropriate to calibrate empirical data such as water and sediment.

Section 5.4.: Why aren’t the results presented in terms of dioxins and furans (e.g. their surrogate 2,3,7,8-PeCDF)?

Section 5.4.3.2, Model Predictions Compared to Individual Samples: These graphs would be more helpful if they weren’t on a log scale. The scale hides some of the more extreme outliers, but it is still apparent that there is quite a range in concentration.

Figure 5-18: The model does not seem to predict the spikes in concentration around the International Slip , RM 7E, 11E and Swan Island. There is likely a fundamental problem on how sediment SWACs are representing exposure for smallmouth bass tissue. As shown in Figures 1 and 2, smallmouth bass respond quickly to changes in sediment concentration. Tightning up the sediment exposure area to only include near shore sediment covered by the composite (instead of in some cases taking 3 miles), may improve model performance.

Figure 5-19, PCB 77 and Smallmouth Bass: This figure illustrates one of the bigger shortcomings of the model, which is its ability to predict dioxin like congeners and segments of river elevated. This may be due to the inclusion of metabolic factors, or another model incongruity. In either case, as shown by Figures 3 and 4 in these comments, River Mile 2E is one of the most significant sources of dioxin like risk – especially PCB 77. However, these important empirical results are not replicated by the model. Total PCB concentration may underestimate risk from PCB dioxin like PCB in this case. Other areas where this occurs in SB 7W (ARKEMA area with elevated PCB 77). More work should be completed to identify the source of this error.

Figure 5-20, PCB 126: This congener shows a similar pattern as above, but the Swan Island data point is of particular concern. The effect of detection limits should be discussed relative to PCB 126.

Figures 5-22 and 5-24, Sum DDD, Sum DDT, and Total DDX: The model is under predicting at RM 7W (Round 3 sample on one side of the river). A BSAF approach or localized food web model should be used to better understand bioaccumulation in this area.

Section 5.6.7.1, Influence of Selected Water Concentrations on the PRG: RI data is needed to review this section.

Section 5.6.7.2, Influence of Model Calibration on PRG Estimates: A true measure of sensitivity needs to be calculated calibrated parameters, which was not evaluated (e.g. metabolic rates).

#### **Appendix D:**

Section 2.0: This comment was a previous Bruce H comment and EPA carried it forward in the last iteration – it still very much applies. This model is designed to predict only the average tissue concentration. This assumption has lead to use of spatially-averaged input concentrations and parameterization of some variables with standard errors rather than standard deviations. This will produce distributions that do not necessarily reflect the full range of uncertainty in underlying data. Our expectation for a probabilistic model was that it would allow PRGs to be selected from a full distribution of concentrations, not just those constrained around the mean. The model needs to be re-configured to reflect more than just the mean.

Section 2.0, Number 4: The text states that “for all other parameters with insufficient data to define a distribution”. This is an instance where use of triangular distributions imparts information (how likely it is that a certain food item will be available, caught, and consumed) that is nearly impossible to actually know. Again, parsimony would indicate use of a range (uniform distribution), possibly truncated high or low if there is information regarding how little or how much of a food item is consumed.

Section 3.0, Chemical Data: It seems that the use of chemical data would be perfect to calibrate, unlike actual environmental parameters like water concentration. Perhaps this is where calibration, if any, is focused.

Section 4.1.1, Temperature Data: This issue has been debated before. In the last set of comments, EPA determined that a mean temperature should not be used, since this does not reflect much of the actual

time the river is at a given temperature. The comments stated “surface water temperature ranges from 6 – 23 C in a predictable fashion through the year and has a direct impact on feeding rate and metabolic function. EPA recommends that an upper bound estimate for temperature rather than a mean temperature be considered.”

Section 4.1.2, Water Chemistry Data: Averages from east, middle and west were calculated to represent each “event”. Uncertainty was then represented by the standard error and not the standard deviation. The data was further averaged into a temporal average based on the hydrographs (the high flow average was weighted as 22% of the total, while the low flow/storm water average was weighted as 78% of the total. The applicability of averaging the water data this to the modeling effort is low and high in uncertainty. We don’t even have a good representation of water variability from this effort, since standard errors instead of standard deviations were used (standard errors were further “weighted” based on hydrograph information). The text states “because of the high level of uncertainty in the high-flow and low flow averages, the uncertainty in the weighting values was not considered when calculated the distribution to be used during model calibration”. This makes no sense to me (high uncertainty is exactly why we need it), and we won’t get more information until the DRAFT RI/FS is submitted (see text).

Section 4.2.3, Chemical Concentrations in Sediment: Standard deviations should be presented, and used in place of the standard error in the model.

Sections 5.1-5.3: Previous comments from EPA stated not to calibrate these parameters (A and B, Dietary chemical transfer efficiency, and proportionality constants. This statement does not preclude the use of distributions to represent these variables instead of one point value. Distributions should be used where applicable.

Section 5.4, Section 6.6, and Table 6-5, Metabolism: The addition of metabolism terms from PCB 77, PCB 126, 2,3,7,8-PeCDF and sum DDT are new to this model. Previous versions assumed metabolism was zero because “metabolism (KM) was not included for most chemicals because they are not thought to be heavily metabolized and/or metabolism data were lacking.” It is unclear why they were added here, but the applicability of these rates should be further reviewed. Table 6-5 shows that the LWG assumes there is significant metabolism of PCB 77 and PCB 126. The rationale for this conclusion is “low concentrations of PCB 77 and PCB 126 in Willamette River tissue, moderate chlorination, so likely to be metabolized by fish”. There is no reference to support this conclusion, and the reference to patterns in Willamette River fish can certainly be explained by other uncertainties and use of the data. Based on the text on Page 29, metabolic rates (Km) were also applied to invertebrates, and calibrated to improve model predictions of empirical data. Metabolic rates should not be calibrated, and it is my opinion that they shouldn’t have been included.

Section 6.7, Dietary Assumptions: Sculpin: This version of the model did not include consumption of juvenile fish by sculpin, which likely underestimates the sculpin ingestion. Largescale Sucker: Zooplankton and phytoplankton should be removed as dietary fractions of the diet. Carp and Northern Pike/minnow: Phytoplankton should be removed as dietary fractions of the diet. The removal of these

fractions are especially necessary given “sediment and tissue consumption was determined as a percentage of the species’ overall diet. The model is set up to normalize these dietary fractions to ensure they always equal 100%”. It is not logical to say, for example, that 45% of a carps diet is phytoplankton and that concentrations represented by phytoplankton ingestion predict carp tissue concentrations.

**Appendix F, Round 3 Data Compared to the Round 2 Report Mechanistic Model:** This section does not present a comparison for dioxins and furans, which were very different in Round 2 data. These SPAFs must be based on an average of the Round 2 compared with an average of Round 3. Why aren’t they compared on a sample by sample basis, where the SPAF a measure of variability in the predictions. How do you model based on an average with a dioxin furan spike at RM 7 like is shown in figures 1 and 2?

Analytical Differences: There are several statements that show different concentrations in Round 3 fish as compared to Round 1. The difference is attributed to the use of high resolution analytical method to reduce interference and detection limits in Round 3, whereas in Round 1 and in most of the sediment sampling SW8081/8270 was used. The use of high res tissue data with the 8081 data could present many problems for food web and BSAF / R modeling. This should be discussed.

**Appendix G, Empirical Tissue Concentrations for the Mechanistic Model:**

Table 2, Tissue Used in the Model: Why weren’t mussels used in the model?

For clams, I show 68 clam samples in Query Manager, but only 38-41 show up in this table? Which stations were not used and why? For crayfish, only 17 are shown for total PCBs, but 37 are showing in Query Manager. It looks like many of the clam and crayfish stations were not included in the model. Submittal of the database used for the modeling effort should be submitted to further answer these questions.

Figure 1: Smallmouth Bass dioxin and furan and dioxin like PCB (pg/g) distribution and contribution to mammalian TEQ by river mile. To calculate a total TEQ the values for the individual congeners can be added for each river mile.

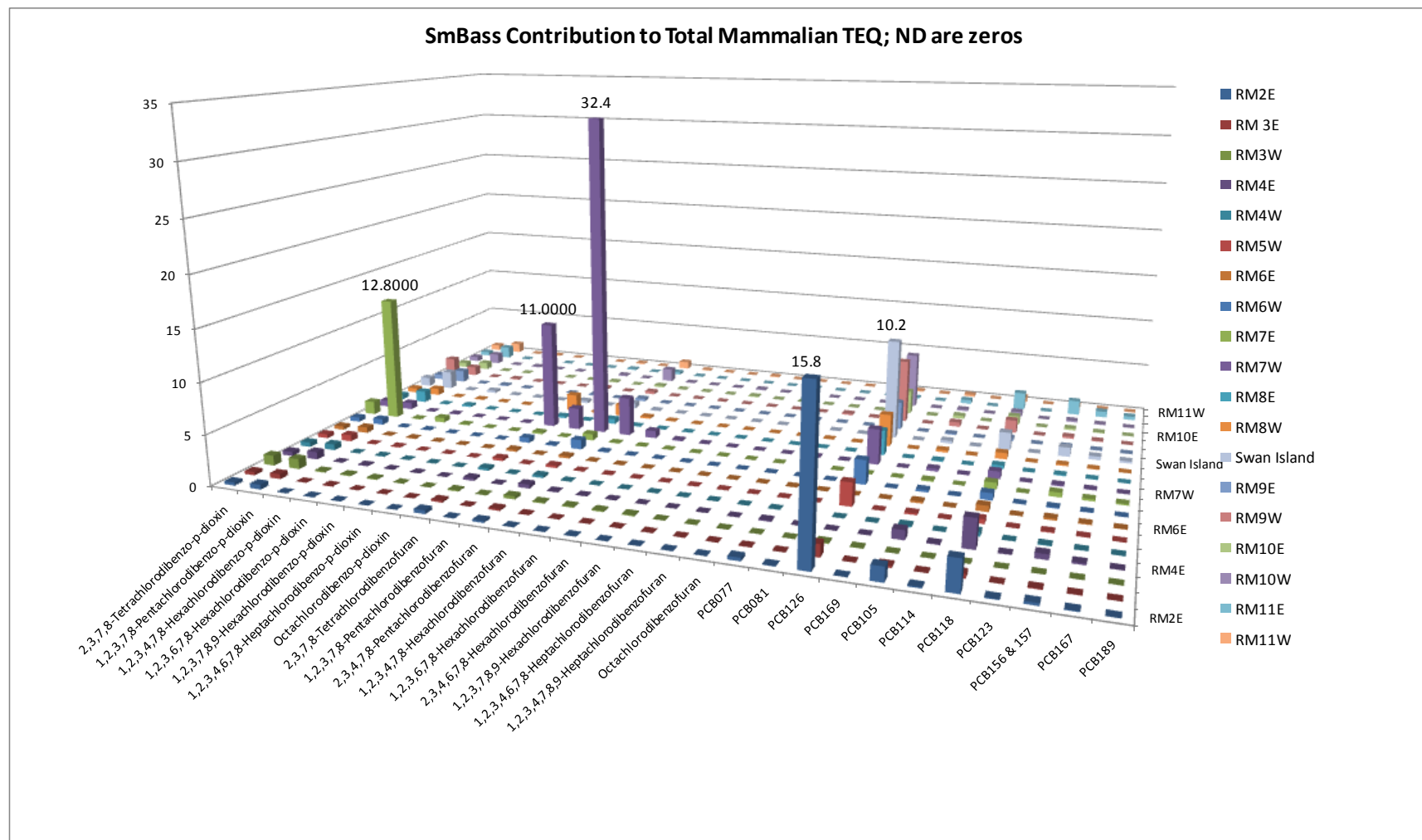


Figure 2, Smallmouth Bass dioxin and furan and dioxin like PCB (pg/g) distribution and contribution to bird TEQ by river mile. \*\* Scale is different than Figure 1. To calculate a total TEQ the values for the individual congeners can be added for each river mile.

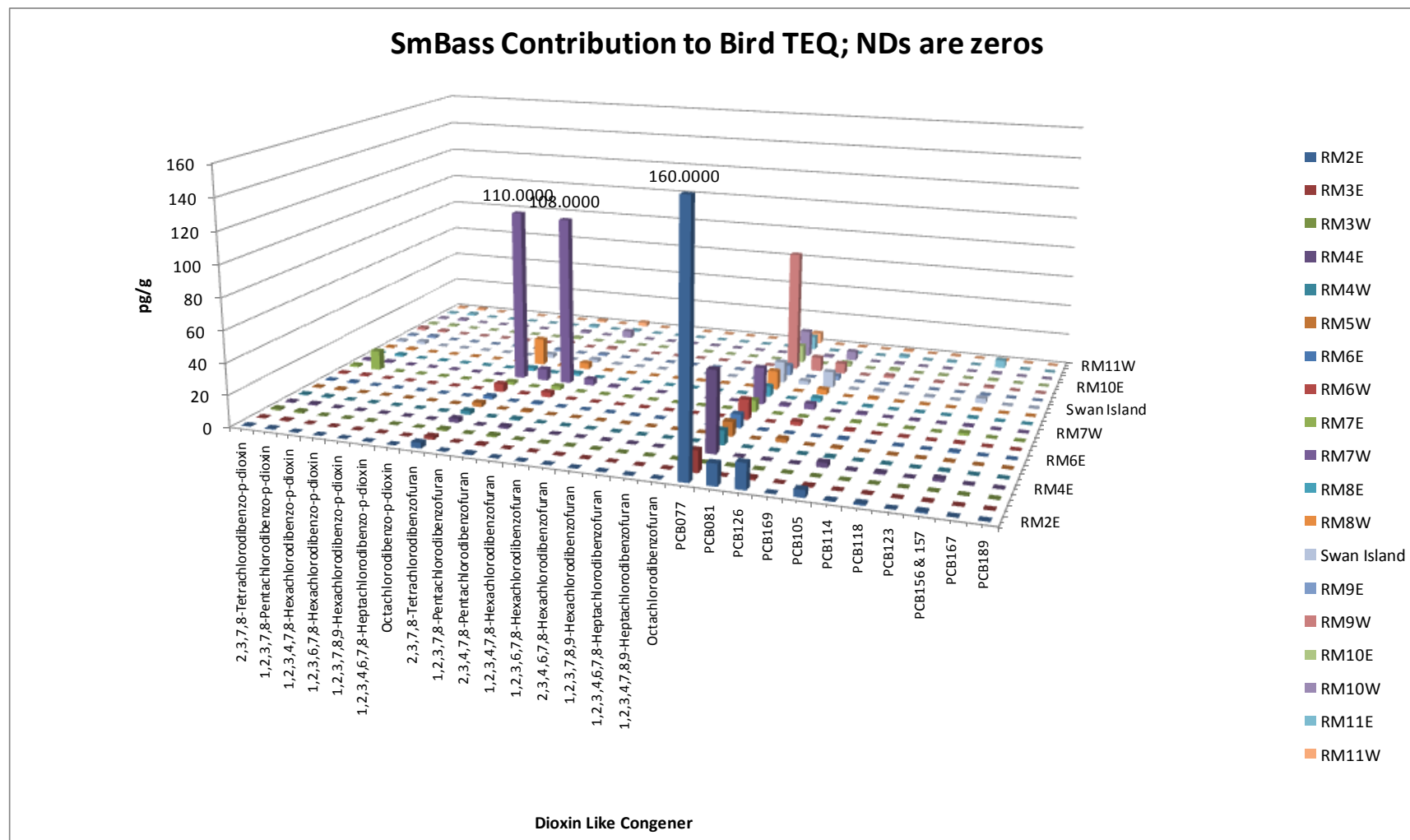


Figure 3: Total PCB concentrations (ppb, y-axis) relative to dioxin TEQ (ppb, x-axis) in Smallmouth Bass tissue. This indicates where Total PCB may not be a good surrogate for PCB dioxin TEQ risk (e.g. sample SB02E). \*\*RM 11E confounded by elevated detection limits (this analysis based on ND = zero)

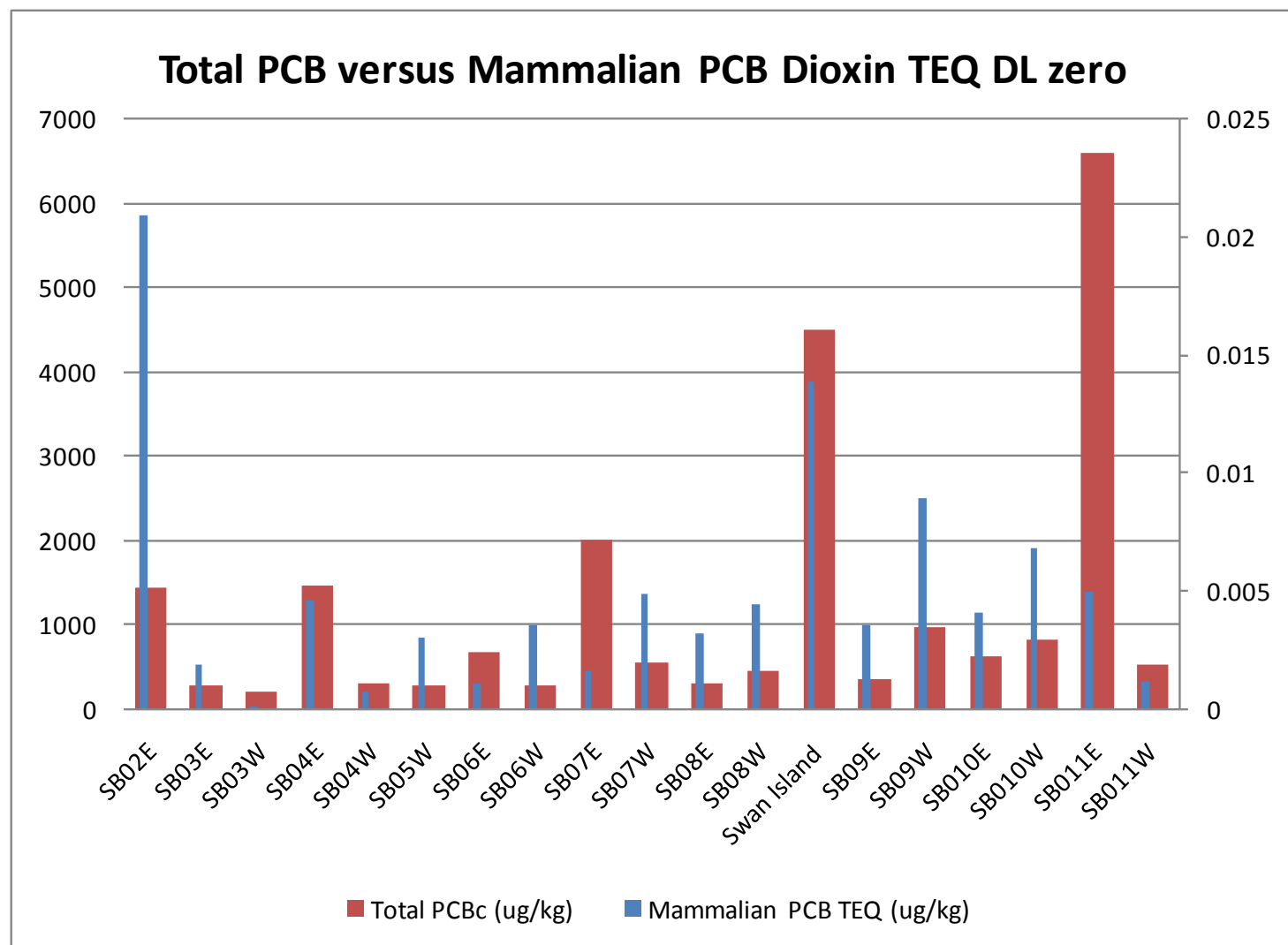


Figure 4: Total PCB (ppb, y-axis) as relative to bird PCB dioxin TEQ (pbb, x-axis). This indicates where Total PCB may not be a good surrogate for dioxin TEQ risk (e.g. sample SB02E). \*\*RM 11E confounded by elevated detection limits (this analysis based on ND=0)

